Lecture 2 Follow up

Sampling Distribution of the Mean

When a sample or a set of samples is drawn repeatedly and randomly from a statistical population, and the mean of each sample alongside its corresponding probability is calculated, the collection of these results is known as the sampling distribution of the mean. The sampling distribution of the mean also has its own mean (denoted as $u_{\bar{x}}$and standard deviation $\sigma_{\bar{x}}$)and in this context, there are two fundamental theorems that illustrate the relationship between the sampling distribution of the mean and the original population.

Theorem 1:

If repeated samples of size (n) are drawn from a population of size (N), then:

For both sampling with replacement and without replacement:

$$u_{\bar{x}} = u_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad or \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Equa1: with replacement

Equa2: without replacement

The second formula(Equa1) is used for finite populations or generally when the following condition is met:

$$n \geq 0.05N$$

).

Theorem 2: The Central Limit Theorem (Normal Distribution)

The normal distribution is one of the most pervasive and renowned distributions, owing to its extensive and varied applications. The distribution is termed "normal" because it represents the regular or normal cases. As the sample size increases .

$$(n \to \infty)$$

the sampling distribution of the mean approximates the normal distribution, regardless of the original population's distribution shape. This approximation happens when the sample size is large ((n \geq 30)), irrespective of the original population's distribution.

When the sampling distribution of the mean follows a normal distribution, the values of the original population elements (x_i) can be transformed from raw or non-standardized values to standardized values, each (x_i) corresponding to a standardized value (z_{x_i}) through the following formula:

$$\mathrm{z}_{x_i} = \frac{x_i - u_{\bar{x}}}{\sigma_{\bar{x}}}$$

Using the standard normal distribution tables and the formula above, we can calculate probabilities pertaining to the population values or the sampling distribution values.

Example:

Consider a statistical population consisting of 925 products from a factory with a mean of 25 units and a standard deviation of 15 units. A random sample of 45 products is drawn.

The task is to:

Calculate the mean of the sampling distribution.

Calculate the standard deviation of the sampling distribution of the mean.

If the sample size becomes 64, calculate $\sigma_{\bar{x}}$ in this case.

Calculate the probability that the mean of a random sample is greater than or equal to 27.5 units.

Solution:

Given (\$u_x = 25$) units,

$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \Rightarrow \sigma_{\bar{x}} = \frac{15}{\sqrt{25}} = 2.2360$ units,

Calculating ($\sigma_{\bar{x}}$) for (n = 64),

Given (n = 64),

Then, as per condition ($n \geq 0.05N$),

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{15}{8}\sqrt{\frac{861}{924}} = 1.875\sqrt{0.9318} = 1.80975$$

Calculating the probability $P(\bar{x} \geq 27.50)$

Since ($30 \leq n$), the sampling distribution of $\bar{x}$follows a standardized normal distribution, where:

$$Z = \frac{\bar{x} - u_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$Z_{27.5} = \frac{27.5 - 25}{2.2360}$$

$$= \frac{2.5}{2.2360}$$

$$= 1.11$$

$$\langle=\rangle P(Z \geq 1.11)$$

This is the probability that the mean of a random sample is greater or equal to 27.5 units.

- **Criteria for Using the Normal Distribution: (Conditions under which the sampling distribution of the mean follows a normal distribution):**
  - **We discuss two cases:**

The first case (known as the Central Limit Theorem): If the sample size is large $n \geq 30$, then the sampling distribution of the mean follows a normal distribution, regardless of the original population's distribution shape.

The second case: If the original population's distribution is normal or nearly normal and the population standard deviation is known, then the sampling distribution of the mean follows a normal distribution.

- **Sampling Distribution of the Mean Using the t-Distribution:**

According to the Student's t-distribution or simply, the t-distribution theory, if the original population follows or approximates a normal distribution, and the population standard deviation is unknown, and the sample size is significantly less than 30 ((n < 30)), then, under these conditions, the sampling distribution of $\bar{x}$follows the t-distribution with (v = n-1) degrees of freedom, where (v) or (df) represents degrees of freedom according to the references used. The t-distribution is similar to the normal distribution but the normal distribution is more "spread out" or "flatter" than the t-distribution.